

# Probabilidade e Estatística

Trabalhando com Pequenas Amostras: Distribuição  $t$  de Student



# Pequenas amostras x Grandes amostras

- Nos exemplos tratados até agora:
  - amostras grandes ( $n > 30$ )
  - qualquer tipo de distribuição original da variável aleatória possui distribuição das médias “adequadamente aproximadas” a uma distribuição normal.
    - Teorema Central do Limite
- Tempo e custo impõem limites ao tamanho da amostra
- Utilização da distribuição normal inadequada para amostras pequenas

# Estimativa da média para pequenas amostras

- Para casos onde  $n \leq 30$  e:
  - População original tem distribuição normal; e
  - O valor de  $\sigma$  é conhecido, podemos calcular E com:
- Nos casos onde  $n \leq 30$  e:
  - População original tem distribuição normal; e
  - O valor de  $\sigma$  é desconhecido, devemos calcular E com a **Distribuição 't' de Student.**

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

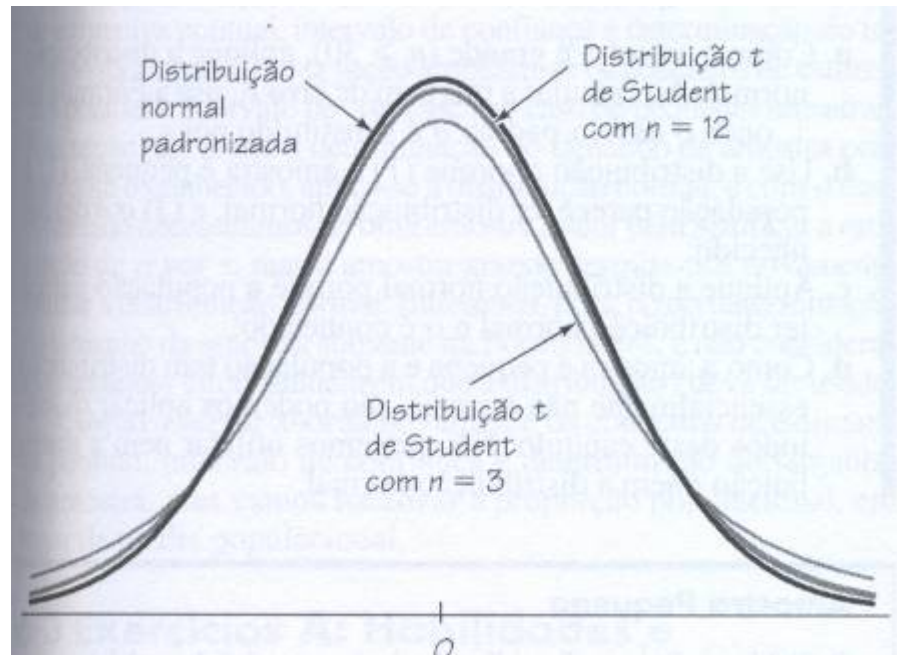
# Distribuição 't' de Student

- Criada por William Gosset (1876-1937)
  - Análises com pequenas amostras
- Empregado da Cervejaria Guinness
- Cervejaria não permitia publicação de pesquisas
  - Pseudônimo de *Student*

# Distribuição 't' de Student

- A **Distribuição 't' de Student** é essencialmente uma distribuição normal (com forma aproximada de um sino) para todas as amostras de tamanho 'n'.
- Através dela, determinamos os valores críticos  $t_{\alpha/2}$  do intervalo de confiança onde

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$



# Utilização da Tabela

- Obtemos o valor de  $t_{\alpha/2}$  na Tabela A3 localizando o número de graus de liberdade na coluna à esquerda e percorrendo a linha correspondente até atingir o número diretamente abaixo do valor aplicável (bilateral) de  $\alpha$ .
- Grau de liberdade: para um conjunto de dados correspondente ao número de valores que podem variar após terem sido impostas certas restrições a todos os valores.
  - Ex: 10 estudantes obtêm em um teste média 8,0
  - A soma das 10 notas deve ser 80. Portanto, se temos um grau de liberdade de  $10-1=9$ , as nove primeiras notas podem ser escolhidas aleatoriamente, contudo a 10ª deve ser igual a  $[80-(\text{soma das 9 primeiras})]$ .
- Aplicaremos, em nosso curso,  $(n-1)$  graus de liberdade

# Propriedades da Distribuição t de Student

- É diferente conforme o tamanho da amostra ( $n$ )
- Tem forma geral simétrica, mas reflete a maior variabilidade esperada em pequenas amostras
- Tem média  $t=0$
- O desvio padrão varia com o tamanho da amostra, mas é superior a 1
- Quanto maior 'n', maior a aproximação em relação à distribuição normal. Para  $n>30$  podemos utilizar distribuição normal com valores críticos 'z'.
- Condições de utilização
  - Tamanho da amostra pequeno ( $n \leq 30$ )
  - $\sigma$  desconhecido
  - População original tem distribuição essencialmente normal

# Assim, para...

- Amostras com as condições anteriores:

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Onde  $t_{\alpha/2}$  tem  $n-1$  graus de liberdade

- Intervalo de confiança

$$\bar{x} - E < \mu < \bar{x} + E$$

ou

$$\mu = \bar{x} \pm E$$

$$(\bar{x} - E; \bar{x} + E)$$



# Observações finais

- Para aplicarmos t de Student, a distribuição da população original deve ser essencialmente normal.
- Contudo, pode-se obter bons resultados se:
  - For basicamente simétrica;
  - Possuir uma única moda.
- Aplicamos t de Student no lugar da distribuição normal, pois quando não se conhece  $\sigma$ , a utilização de 's' de uma pequena amostra incorpora outra fonte de erro. Para mantermos o grau de confiança desejado, compensamos a variabilidade adicional ampliando o intervalo de confiança por um processo que substitui o valor crítico  $z_{\alpha/2}$  por outro  $t_{\alpha/2}$  obtido na tabela da Distribuição t de Student.

# Exemplo

- Considere um teste de colisão de carros. A análise de 12 carros danificados resulta num custo de conserto que parece ter distribuição em forma de sino, com média e desvio-padrão a seguir (R\$).

$$\bar{x} = 26.227$$

$$s = 15.873$$

- Determine:
  - a melhor estimativa pontual de  $\mu$  (custo do conserto)
  - O intervalo de confiança para NC=95%

# Exemplo

Obs.: valores em R\$

a)  $\bar{x} = 26.227$

b) Amostra pequena ( $n \leq 30$ ); desvio padrão desconhecido; distribuição é similar à distribuição normal

Na tabela: para a coluna 0,05 bilateral e grau de liberdade  $n-1=11 \rightarrow t_{\alpha/2}=2,201$

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 2,201 \cdot \frac{15873}{\sqrt{12}} = 10.085,29$$

$$\bar{x} - E < \mu < \bar{x} + E$$

$$26.227 - 10.085,29 < \mu < 26.227 + 10.085,29$$

$$16.141,71 < \mu < 36.312,29$$

$$\mu = 26.227 \pm 10.085,29$$

Existe uma probabilidade de 95% do intervalo de confiança conter efetivamente a média da população: custos de reparo

# Síntese da Estimativa de Média Populacional

